AD 607350

D1-82-0371

BOEING SCIENTIFIC RESEARCH LABORATORIES

Estimation of The Transition Distributions of a Markov Renewal Process

Erin H. Moore

Ronald Pyke

Mathematics Research

September 1964

BLANK PAGE

D1-82-0371

ESTIMATION OF THE TRANSITION DISTRIBUTIONS OF A MARKOV RENEWAL PROCESS

by

Erin H. Moore
Airplane Division, The Boeing Company

and

Ronald Pyke
University of Washington

Mathematical Note No. 364

Mathematics Research Laboratory

BOEING SCIENTIFIC RESEARCH LABORATORIES

September 1964

ABSTRACT

The present paper is concerned with the estimation of the transition distributions of a Markov renewal process with finitely many states, which extends and unifies some aspects of the results in the special cases of discrete and continuous parameter Markov chains. A natural estimator of the transition distributions is defined and shown to be consistent. Limiting distributions of this estimator are derived. A density for a Markov renewal process is developed to permit the definition of maximum likelihood estimators for a renewal process and for a Markov renewal process.

BLANK PAGE

INTRODUCTION

The general theory of statistical inference in Markov processes began with Bartlett's paper in 1951, [2]. Later developments are presented in Billingsley's book [4] and his expository paper [5], both of which appeared in 1961. We refer in particular to the development of maximum likelihood estimators for the transition probabilities of a Markov chain, either discrete or continuous parameter, by Billingsley [4] and more recently by Albert [1] in 1962. The present paper is concerned with the estimation of the transition distributions of a Markov renewal process with finitely many states, which extends and unifies some aspects of the results in the special cases of discrete and continuous parameter Markov chains. In Chapter 2 a natural estimator of the transition distributions is defined and shown to be consistent. Limiting distributions of this estimator are derived in Chapter 3. A density for a Markov renewal process is developed in Chapter 4 to permit the definition of maximum likelihood estimators for a renewal process in Chapter 5 and for a Markov renewal process in Chapter 6.

1. PRELIMINARY CONCEPTS AND DEFINITIONS

The constructive definition given in [11] of a Markov renewal process (MRP) with m $(< \infty)$ states is briefly as follows. One is given a matrix of transition distributions (Q_{ij}) where each Q_{ij} is a mass function defined on $(-\infty,\infty)$ satisfying $Q_{ij}(x) = 0$ for $x \leq 0$ and $\sum_{j=1}^{m} Q_{ij}(\infty) = 1$, $(1 \leq i \leq m)$. One is also given an m-tuple of initial probabilities $(p_1, p_2, \cdots, p_m)$ which satisfies $p_j \geq 0$ and $\sum_{j=1}^{m} p_j = 1$. Consider any two-dimensional Markov process $\{(J_n, X_n); n \geq 0\}$ defined on a complete probability space that satisfies $X_o = 0$ (a.s.), $P[J_o = k] = p_k$ and

$$P[J_n = k, X_n \leq x | J_o, J_1, \cdots, J_{n-1}, X_1, \cdots X_{n-1}] = Q_{J_{n-1},k}(x) \qquad \text{(a.s.)}$$

for all $x \in (-\infty, \infty)$ and $1 \leq k \leq m$. The matrix (p_{ij}) is defined by $p_{ij} = Q_{ij}(\infty)$. If $p_{ij} > 0$, set $F_{ij} = p_{ij}^{-1} Q_{ij}$, while if $p_{ij} = 0$, then let F_{ij} be arbitrary. The integer-valued stochastic processes $\{N(t); t \geq 0\}$, $\{N_j(t); t \geq 0\}$, and $\{N_{ij}(t); t \geq 0\}$ are defined by $N(t) = \sup \{n \geq 0 : \sum_{i=0}^{n} X_i \leq t\}$, $N_j(t)$ = the number of times $J_k = j$ for $1 \leq k \leq N(t)$, and $N_{ij}(t)$ = the number of times $J_k = i$ and $J_{k+1} = j$ for $1 \leq k \leq N(t)$. Then the stochastic process $\{N_1(t), N_2(t), \cdots, N_m(t); t \geq 0\}$ is called a Markov renewal process determined by the given initial probabilities and matrix of transition distributions.

The following consequences of the above definitions, derived in [11], will be used below.

$$(1.1) \quad \begin{cases} P[J_n = j | J_o, \cdots, J_{n-2}, J_{n-1} = i] = p_{ij} \\[2ex] P[X_n \leq x | J_o, \cdots, J_{n-2}, J_{n-1} = i, J_n = j] = F_{ij}(x) \\[2ex] P[X_{n_1} \leq x_1, \cdots, X_{n_k} \leq x_k | J_n; n \geq 0] = \prod_{i=1}^{k} F_{J_{n_i - 1}, J_{n_i}}(x_i) \end{cases}$$

for $0 < n_1 < \cdots < n_k$, the last equality holding with probability one.

It is assumed throughout that the MRP is irreducible, recurrent, and that $F_{ij} = H_i$ for $1 \leq j \leq m$. This last assumption incurs no loss of generality as is pointed out in [12].

Estimators for the transition probabilities $Q_{ij}(x)$ are defined on sample functions of the MRP over $[0,t]$. These sample functions of the MRP are equivalent to the sample functions $(J_o, J_1, \cdots, J_{N(t)}, X_1, X_2, \cdots, X_{N(t)})$. Let X_{ij} denote the holding time of the j^{th} visit to state i, that is the $\{X_{ij};\ 1 \leq i \leq m,\ 1 \leq j \leq N_i(t)\}$ are just a relabeling of $\{X_i;\ 1 \leq i \leq N(t)\}$.

2. DEFINITION AND CONSISTENCY OF A NATURAL ESTIMATOR

Consider the estimator defined by

$$\hat{Q}_{ij}(x;t) = \hat{p}_{ij}(t)\, \hat{H}_i(x;t), \tag{2.1}$$

where $t, x > 0$,

$$\hat{p}_{ij}(t) = N_{ij}(t)/N_i(t), \tag{2.2}$$

$$\hat{H}_i(x;t) = N_i(t)^{-1} \sum_{k=1}^{N_i(t)} \epsilon(x - X_{ik}), \tag{2.3}$$

and where $\epsilon(u)$ equals one if $u \geq 0$ and zero otherwise. That is, $\hat{H}_i(x;t)$ is the ordinary empirical distribution function but determined from the sample, of random size $N_i(t)$, of the holding times in state i. Interpret $\hat{Q}_{ij}(x;t)$ to be zero if $N_i(t) = 0$.

The estimator (2.1) is a natural combination of estimators used in Markov chain inference and in classical inference for fixed sample size. Derman [8] has studied $\hat{p}_{ij}(t)$ as an estimator for the transition probabilities of a Markov chain, with the small difference that the total number of transitions, $N(t)$, is not random. The empirical distribution function for non-random sample size has been studied extensively (c.f. Darling, [7]).

Consistency of (2.1) and the limiting distributions of (2.1) are obtained using the general limit theorems for MRP developed by Pyke [12]. In [12] the limiting behavior (as $t \to \infty$) of sums of the form

$$W_f(t) = \sum_{n=1}^{N(t)} f(J_{n-1}, J_n, X_n) \tag{2.4}$$

is studied for real valued functions f defined on the state space of an MRP. We recall the notation used in [12]. μ_{ij} and μ_{ij}^* denote the first moment of the distribution of the first passage time from state i to state j of the MRP and of the corresponding Markov chain $\{J_n: n \geq 0\}$, respectively. Define recurrence indices $r_{j,s}$ by $r_{j,0} = 0$ and, for $s \geq 1$,

$$r_{j,s} = \sup\{1 \leq k \leq \infty;\ k > r_{j,s-1},\ J_i \neq j\ (r_{j,s-1} < i < k)\}.$$

The sequence of random variables (r.v.'s) $\{U_{j,s}: s > 0\}$ is defined by

$$U_{j,s} = \sum_{n=r_{j,s}+1}^{r_{j,s+1}} f(J_{n-1}, J_n, X_n). \tag{2.5}$$

That is, $U_{j,s}$ is the contribution to the sum $W_f(t)$ obtained between the s^{th} and the $(s+1)^{th}$ occurrence time of state j. The random variables $\{U_{j,s};\ s \geq 1\}$ are independent and identically distributed. Set

$$A_{ik} = \int_0^\infty f(i,k,x)\, dQ_{ik}(x), \quad A_i = \sum_{k=1}^m A_{ik}$$

$$B_{ik} = \int_0^\infty [f(i,k,x)]^2 dQ_{ik}(x), \quad B_i = \sum_{k=1}^m B_{ik}.$$

When the mean and variance of $U_{i,1}$ exist, they will be denoted by m_i, and σ_i^2 respectively. Since m is finite, it follows from [12] that when they exist, they are given by

$$m_i = \sum_{r=1}^m A_r \mu^*_{ii} / \mu^*_{rr} \tag{2.6}$$

and

$$\sigma_i^2 = -m_i^2 + \sum_{r=1}^m B_r \mu^*_{ii} / \mu^*_{rr} \tag{2.7}$$

$$+ 2 \sum_{r=1}^m \sum_{s \neq i} \sum_{k \neq i} A_{rs} A_k \mu^*_{ii} (\mu^*_{si} + \mu^*_{ik} - \mu^*_{sk}) / \mu^*_{rr} \mu^*_{kk}.$$

Theorem 2.1: The estimator (2.1) is uniformly strongly consistent as $t \to \infty$ in the sense that with probability one,

$$\max_{i,j} \sup_x |\hat{Q}_{ij}(x;t) - Q_{ij}(x)| \to 0. \tag{2.8}$$

Proof: Rewrite (2.8) as

$$\max_{i,j} \sup_x |[N_{ij}(t)/N_i(t) - p_{ij}]\, \hat{H}_i(x;t) + p_{ij}[\hat{H}_i(x;t) - H_i(x)]|$$

$$\leq \max_{i,j} |N_{ij}(t)/N_i(t) - p_{ij}| + \max_i \sup_x |\hat{H}_i(x;t) - H_i(x)|.$$

Since $N_i(t) \to \infty$ (a.s.) by the regularity of the MRP, then one concludes from the Glivenko-Cantelli theorem for non-random sample sizes, that $\sup_x |\hat{H}_i(x;t) - H_i(x)| \to 0$ (a.s.). The proof is completed by showing $[N_{ij}(t)/N_i(t) - p_{ij}] \to 0$ (a.s.) for $1 \leq i, j \leq m$. Let k_ℓ denote the state visited after the ℓ^{th} visit to state i. Then

$$\sum_{\ell=1}^{N_i(t)-1} \delta_{k_\ell,j} \leq N_{ij}(t) \leq \sum_{\ell=1}^{N_i(t)} \delta_{k_\ell,j} \tag{2.9}$$

where $\delta_{k,j}$ denotes the Kronecker delta and by the Strong Law of Large Numbers both the right and left hand sides of (2.9), when divided by $N_i(t)$, converge to p_{ij} with probability one.

3. ASYMPTOTIC DISTRIBUTION OF THE NATURAL ESTIMATOR

The limiting distribution of (2.1), (2.2), (2.3) can be obtained by applying the central limit theorem for functions on an MRP (c.f. Lemma 7.1, [12]).

<u>Theorem 3.1</u>: For fixed i,j,x, $(t^{\frac{1}{2}}[\hat{p}_{ij}(t) - p_{ij}],\ t^{\frac{1}{2}}[\hat{H}_i(x;t) - H_i(x)])$ converges in law as $t \to \infty$ to a bivariate normal r.v. with means zero and covariance matrix (σ_{ij}) given by

$$(3.1)\qquad \sigma_{11} = \mu_{ii}p_{ij}(1-p_{ij}),\ \sigma_{22} = \mu_{ii}H_i(x)[1-H_i(x)],\ \sigma_{12} = \sigma_{21} = 0\,.$$

<u>Proof</u>: Let w_1 and w_2 be arbitrary constants. To prove the asymptotic joint normality it suffices to show that

$$(3.2)\qquad w_1 t^{\frac{1}{2}}[\hat{p}_{ij}(t) - p_{ij}] + w_2 t^{\frac{1}{2}}[\hat{H}_i(x;t) - H_i(x)]$$

converges in law to a normal r.v. for all w_1 and w_2. We rewrite (3.2) as the product of $[t/N_i(t)]$ and a sum of the form (2.4) by using the function f defined by

$$(3.3)\qquad f(r,s,y) = \{w_1[\delta_{sj} - p_{ij}] + w_2[\epsilon(x-y) - H_i(x)]\}\delta_{ri}\,.$$

For the function (3.3)

$$A_r = w_1\delta_{ri}[p_{rj} - p_{ij}] + w_2\delta_{ri}[H_r(x) - H_i(x)] = 0$$

and

$$B_r = \{w_1^2[p_{rj} + p_{ij}^2 - 2p_{ij}p_{rj}] + w_2^2[H_r(x) + H_i^2(x) - 2H_r(x)H_i(x)]\}\delta_{ri}$$

for $1 \leq r \leq m$; hence $m_i = 0$ and the third sum in (2.7) is zero. Then the variance of $U_{i,1}$ is

$$\sigma_i^2 = \sum_{r=1}^{m} B_r\,\mu_{ii}^*/\mu_{rr}^* = w_1^2 p_{ij}[1 - p_{ij}] + w_2^2 H_i(x)[1 - H_i(x)].$$

The variance σ_i^2 is finite, so from Lemma 7.1 of [12] the limiting distribution of $t^{-\frac{1}{2}}W_f(t)$ for the f given in (3.3) is normal with zero mean and variance σ_i^2/μ_{ii}. But $t/N_i(t) \to \mu_{ii}$(a.s.) so the limiting distribution of (3.2) is normal with zero mean and variance $\mu_{ii}\sigma_i^2$ as required.

The zero correlation between $\hat{p}_{ij}(t)$ and $\hat{H}_i(x;t)$ yields the following result.

<u>Corollary 3.2</u>: For fixed i,j,s, $\hat{p}_{ij}(t)$ and $\hat{H}_i(x;t)$ are asymptotically independent.

The asymptotic normality of (3.2) can be used to obtain the limiting distribution of $\hat{Q}_{ij}(x;t)$.

<u>Corollary 3.3</u>: For fixed i,j,x, $t^{\frac{1}{2}}[\hat{Q}_{ij}(x;t) - Q_{ij}(x)]$ converges in law as $t \to \infty$ to a normally distributed r.v. with mean zero and

variance equal to

(3.4) $$\mu_{ii}H_i(x)p_{ij}[H_i(x) - 2H_i(x)p_{ij} + p_{ij}].$$

<u>Proof</u>: We rewrite $t^{\frac{1}{2}}[\hat{Q}_{ij}(x) - Q_{ij}(x)]$ as

(3.5) $$t^{\frac{1}{2}}\hat{H}_i(x;t)[\hat{p}_{ij}(t) - p_{ij}] + t^{\frac{1}{2}}p_{ij}[\hat{H}_i(x;t) - H_i(x)].$$

By a well known convergence theorem (Cramer [6], Section 20.6) the limiting distribution of (3.5) is the same as the limiting distribution of

(3.6) $$t^{\frac{1}{2}}H_i(x)[\hat{p}_{ij}(t) - p_{ij}] + t^{\frac{1}{2}}p_{ij}[\hat{H}_i(x;t) - H_i(x)].$$

With the particular choice $w_1 = H_i(x)$ and $w_2 = p_{ij}$, (3.2) is just (3.6) and the proof is complete.

The asymptotic normality given in Corollary 3.3 can be extended to the finite dimensional distribution of the r.v.'s $\{W_{ijk} = \hat{Q}_{ij}(x_k;t) - Q_{ij}(x_k)$ for $1 \leq i,\ j \leq m$ and $1 \leq k \leq s\}$.

<u>Theorem 3.4</u>: For fixed s, the distribution of $\{t^{\frac{1}{2}}W_{ijk};\ 1 \leq i,\ j \leq m,\ 1 \leq k \leq s\}$ converges in law as $t \to \infty$ to an m^2s-dimensional normal r.v. with zero mean and covariance matrix $(a_{ijk,uvw})$ given by

(3.8) $$a_{ijk,uvw} = \mu_{ii}\delta_{iu}p_{ij}[H_i(x_w)\delta_{jv} + H_i(\min[x_k,x_w])p_{iv} - 2H_i(x_k)H_i(x_w)p_{iv}].$$

Proof: Let $\{\lambda_{ijk};\ 1 \le i,\ j \le m,\ 1 \le k \le s\}$ be arbitrary constants. It will suffice to show that

$$(3.9) \qquad t^{\frac{1}{2}} \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{k=1}^{s} \lambda_{ijk} W_{ijk}$$

converges in law to a normal r.v. for all real λ_{ijk}. We may rewrite (3.9) as

$$[t/N_1(t)]\ t^{-\frac{1}{2}} \sum_{i=1}^{m} [N_1(t)/N_i(t)] \sum_{j=1}^{m} \sum_{k=1}^{s} \lambda_{ijk}$$

$$\cdot([N_{ij}(t) - p_{ij}N_i(t)]\hat{H}_i(x_k;t) + p_{ij}N_i(t)[\hat{H}_i(x_k;t) - H_i(x_k)]).$$

As in the proof of Theorem 3.1, the expression may be shown to have the same limiting distribution as

$$[t/N_1(t)]t^{-\frac{1}{2}} \sum_{i=1}^{m} [\mu_{ii}/\mu_{11}] \sum_{j=1}^{m} \sum_{k=1}^{s} \lambda_{ijk}$$

$$\cdot[N_{ij}(t)H_i(x_k) + p_{ij}N_i(t)\hat{H}_i(x_k;t) - 2p_{ij}N_i(t)H_i(x_k)].$$

This in turn can be written as a product of $[t/N_1(t)]$ and a sum of the form (2.4) by using the function f defined by

$$(3.10) \qquad f(r,s,y) = \mu_{11}^{-1} \sum_{i=1}^{m} \mu_{ii} \sum_{j=1}^{m} \sum_{k=1}^{s} \lambda_{ijk}\, \delta_{ri}$$

$$\cdot [H_i(x_k)\delta_{sj} + p_{ij}\, \epsilon(x_k - y) - 2H_i(x_k)p_{ij}].$$

For this function,

$$A_r = \mu_{11}^{-1} \sum_{i=1}^{m} \mu_{ii} \sum_{j=1}^{m} \sum_{k=1}^{s} \lambda_{ijk}\, \delta_{ri}$$

$$\cdot [H_i(x_k)p_{rj} + p_{ij}H_r(x_k) - 2H_i(x_k)p_{ij}] = 0$$

for $1 \leq r \leq m$; hence $m_1 = 0$ and the third sum in (2.7) is zero. Then the variance of $U_{1,1}$ is given by

$$\sigma_1^2 = \sum_{r=1}^{m} B_r\, \mu_{11}^{*}/\mu_{rr}^{*}$$

which may be shown to reduce to

$$\sigma_1^2 = \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{k=1}^{s} \sum_{v=1}^{m} \sum_{w=1}^{s} \lambda_{ijk}\, \lambda_{ivw} [\mu_{ii}/\mu_{11}]$$

$$\cdot [H_i(x_k)H_i(x_w)p_{ij}\delta_{jv} - 2H_i(x_k)H_i(x_w)p_{iv}p_{ij}$$

$$+ H_i(\min[x_k, x_w])p_{iv}p_{ij}].$$

The variance σ_1^2 is finite, so by the same argument as in Theorem 3.1, the limiting distribution of (3.9) is normal with zero mean and variance $\sigma_1^2\mu_{11}$. The required covariance matrix (3.8) is obtained from the coefficients of $\lambda_{ijk}\lambda_{uvw}$, thereby completing the proof.

Consider a renewal process, that is, an MRP with one state for which $m = 1$, $p_{11} = 1$, $N_{11}(t) = N_1(t) = N(t)$. From Theorem 3.4 the limiting distribution as $t \to \infty$ of $N(t)^{\frac{1}{2}}[\hat{H}_1(x_k t) - H_1(x_k)]$ for $1 \leq k \leq s$ with s fixed, is normal with zero means and covariance matrix $(a_{k,w})$ defined by

$$a_{k,w} = [H_1(\min[x_k,x_w]) - H_1(x_k)H_1(x_w)].$$

Consider the Markov chain obtained from the MRP by letting the holding times be degenerate at one, that is, $H_i(x) = \epsilon(x-1)$, $\mu_{ii} = \mu_{ii}^*$. From Theorem 3.4 one obtains that as $t \to \infty$, $t^{\frac{1}{2}}[N_{ij}(t)/N_i(t) - p_{ij}]$ for $1 \leq i, j \leq m$ converges to a normal r.v. with zero means and covariance matrix given by

$$a_{ij,uv} = \mu_{ii}^* \delta_{iu} p_{ij}[\delta_{jv} - p_{iv}].$$

This is equivalent to Derman's result on the limiting distribution of $n^{\frac{1}{2}}N_{ij}(n)/N_i(n)$ (c.f. Billingsley [5]).

4. DENSITY FOR A MARKOV RENEWAL PROCESS

A density for an MRP is defined in a manner similar to the definition of a density for a continuous time Markov process by Billingsley [4] and Albert [1]. From the constructive definition of an MRP given in Section 2, almost all sample functions for an MRP up to time t can be represented as the finite tuple $R(t) = (J_o, J_1, \cdots, J_{N(t)}, X_1, \cdots, X_{N(t)})$. Almost every sample function may therefore be represented as a point in $\Omega = \bigcup_{n=o}^{\infty} \Omega_n$ where Ω_n is the $(n+1)$-fold Cartesian product of $\{1,2,\cdots,m\} \times [o,\infty)$. Let $\mathcal{A}_n$ be the product Borel field on Ω generated by all subsets of $\{1,2,\cdots,m\}$ and the ordinary Borel sets on $[o,\infty)$. Let $\mathcal{A}$ be the smallest σ-field containing each $\mathcal{A}_n$, $0 \leq n \leq \infty$. For convenience we will assume that the underlying probability space on which the MRP is defined is $(\Omega,\mathcal{A})$. On this probability space the measure P is as follows.

<u>Theorem 4.1</u>: For any $n \geq 0$ and integers $1 \leq j_i \leq m$ for $0 \leq i \leq n$, the probability measure P on $(\Omega,\mathcal{A})$ is given by

$$(4.1) \qquad P[N(t) = n, J_o = j_o, \cdots, J_n = j_n, X_1 \leq \alpha_1, \cdots, X_n \leq \alpha_n]$$

$$= p_{j_o} \int_{C_n} [1 - H_{j_n}(u_t)] \prod_{k=o}^{n-1} p_{j_k j_{k+1}} \, dH_{j_k}(x_{k+1})$$

where $u_t = t - x_1 - x_2 - \cdots - x_n$ and

$$C_n = \{(x_1, x_2, \cdots, x_n) : 0 \leq x_k \leq \alpha_k, \ 1 \leq k \leq n \text{ and } u_t > 0\}.$$

In particular,

$$P[N(t) = 0, J_o = j_o] = p_{j_o}[1 - H_{j_o}(t)].$$

<u>Proof</u>: From (1.1) the conditional distribution of $\{X_i, 1 \le i \le n\}$ given $\{J_i, 0 \le i \le n\}$ is that of n independent r.v.'s with distribution functions H_{J_i} respectively. The proof follows immediately.

The density of the process can now be exhibited as the Radon-Nikodym derivative of the probability distribution (4.1) with respect to the measure defined as follows. Let μ be Lebesgue measure on $[o,\infty)$, let λ be counting measure on $\{1,2,\dots,m\}$, and let σ_n be the appropriate product measure on $(\Omega_n, \mathcal{A}_n)$. For each set $B \in \mathcal{A}$ define $\sigma^*(B) = \sum_{n=o}^{\infty} \sigma_n(B \cap \Omega_n)$, which determines a measure on $(\Omega, \mathcal{A})$.

The density is now set forth explicitly.

<u>Theorem 4.2</u>: If each H_i is absolutely continuous with density function h_i, then one may write

$$P(B) = \int_B f(v)\, d\,\sigma^*(v), \qquad B \in \mathcal{A},$$

where

$$(4.2)\qquad f(v) = \begin{cases} p_{j_o}[1 - H_{j_o}(t)] & \text{if } v = j_o, \\ p_{j_o}[1 - H_{j_n}(u_t)] \prod_{k=o}^{n-1} p_{j_k j_{k+1}} h_{j_k}(x_{k+1}) & \\ & \text{if } v = (j_o, \cdots, j_n, x_1, \cdots, x_n) \text{ with } u_t > 0 \\ 0 & \text{otherwise} \end{cases}$$

and where $u_t = t - x_1 - x_2 - \cdots - x_n$.

__Proof:__ Let the conditional density of $(J_o, \cdots, J_n, X_1, \cdots, X_n)$ with respect to σ_n, given that $N(t) = n$, be denoted by $g_n(j_o, \cdots, j_n, x_1, \cdots, x_n)$. This conditional density exists since under the stated condition, $R(t)$ is a vector r.v. of fixed dimension whose coordinates are either discrete or absolutely continuous. By Theorem 4.1, $P[N(t) = n]g_n$ must coincide a.e. with f. Thus, for $B \in \mathcal{A}$,

$$P(B) = \sum_{n=o}^{\infty} \int_{B \cap \Omega_n} f d\sigma_n = \int_B f d\sigma^*$$

and so f is the required density with respect to σ^*.

For the special case of exponential holding times (i.e. a continuous Markov process), the density (4.2) reduces to Albert's density (c.f. Theorem 3.2, [1]). For a renewal process, i.e. an MRP with one state, the sample functions are of the form $R(t) = (X_1, X_2, \cdots, X_{N(t)})$ and for $H(x)$ absolutely continuous, (4.2) reduces to

$$(4.3) \qquad f(v) = [1 - H(u_t)] \prod_{i=1}^{n} h(x_i) \quad \text{if} \quad v = (x_1, x_2, \cdots, x_n).$$

5. MAXIMUM LIKELIHOOD ESTIMATION FOR A RENEWAL PROCESS

Maximum likelihood estimators (MLE) may be obtained by maximizing (4.3) over a selected class of densities for an observed sample function $R(t) = (X_1, \cdots, X_{N(t)})$. The classes of densities considered will be exponential, increasing failure rate, and non-increasing. Throughout the remainder of the paper it will be assumed that $H_i(x)$ is absolutely continuous $(1 \leq i \leq m)$ and that whenever t is fixed, $N(t)$ and U_t will be denoted by N and U respectively.

a. Exponential density with parameter λ, that is $h(x) = \lambda \exp(-\lambda x)$.

From (4.3) the likelihood function is

$$L(v) = \exp(-\lambda U) \prod_{k=1}^{N} \lambda \exp(-\lambda X_k)$$

and the log likelihood function is

$$(5.1) \qquad N \log \lambda - \lambda[\sum_{k=1}^{N} X_k + U].$$

The maximum of (5.1) occurs at $\hat{\lambda} = N/t$, so the MLE for $h(x)$ is given by

$$(5.2) \qquad \hat{h}(x) = [N/t] \exp[-Nx/t].$$

The MLE (5.2) is strongly consistent since $N/t \to \lambda$ (a.s.). This example is the well known one of the Poisson process for which the estimator of λ is the same for a fixed-time sample as for a fixed-number-of-events sample.

b. <u>Increasing failure rate</u> (IFR) densities, that is the class of densities for which the failure rate $q(y) = h(y)/[1 - H(y)]$ is increasing. Marshall and Proschan [10] and Grenander [9] have derived the MLE for $q(x)$ based on a sample of non-random size (i.e. $U \equiv 0$ and $N(t) = n$) to be

$$(5.3) \quad \hat{q}(y) = \begin{cases} 0 & \text{for } y < Y_1 \\ \min\limits_{v \geq i+1} \max\limits_{u \leq i} (v-u)[(n-u)(Y_{u+1}-Y_u) + \cdots + (n-v+1)(Y_v-Y_{v-1})]^{-1} & \\ & \text{for } Y_i \leq y < Y_{i+1} (1 \leq i \leq n-1) \\ \infty & \text{for } y \geq Y_n \end{cases}$$

where $\{Y_i;\ 1 \leq i \leq n\}$ are $\{X_i;\ 1 \leq i \leq n\}$ arranged in increasing order. By an argument similar to the one used in [10], the MLE for $q(x)$ can be derived for a renewal process.

<u>Theorem 5.1</u>: Let $(Y_1, Y_2, \cdots, Y_N)$ be an ordered sample from an IFR renewal process. If $Y_{i_o} \le U < Y_{i_o+1}$ for $1 \le i_o \le N-1$ or $U > Y_N$ and $i_o = N$ then the MLE of $q(y)$ is given by

$$(5.4)\quad \hat{q}(y) = \begin{cases} 0 & \text{for } y < Y_1 \\ \min\limits_{v \ge i+1} \max\limits_{u \le i} (v-u)[c_u + \cdots + c_{v-1}]^{-1} & \text{for } Y_i \le y < Y_{i+1} \\ & (1 \le i \le N-1) \\ \infty & \text{for } y \ge Y_N \end{cases}$$

where

$$(5.5)\quad c_i = \begin{cases} (N-i+1)(Y_{i+1} - Y_i) & \text{for } 1 \le i \le i_o \\ (N-i_o)(Y_{i_o+1} - Y_{i_o}) + (U - Y_{i_o}) & \text{for } i = i_o \\ (N-i)(Y_{i+1} - Y_i) & \text{for } i_o < i \le N. \end{cases}$$

If $U < Y_1$, $\hat{q}(y)$ is given by (5.3).

<u>Proof</u>: Since $h = q\exp(-Q)$ and $1 - H = \exp(-Q)$ where $q(y) = h(y)/[1 - H(y)]$ and $Q(y) = \int_0^y q(z)\,dz$, the log likelihood function can be written as

$$(5.6) \qquad \log L = \sum_{i=1}^{N} \log q\,(Y_i) - \sum_{i=1}^{N} Q(Y_i) - Q(U).$$

For $q(y)$ increasing,

$$\sum_{i=1}^{N} Q(Y_i) \geq \sum_{i=1}^{N} (N-i)(Y_{i+1} - Y_i)\, q\,(Y_i)$$

and

$$Q(U) \geq \sum_{i=1}^{i_o-1} (Y_{i+1} - Y_i)\, q\,(Y_i) + (U - Y_{i_o})\, q\,(Y_{i_o}).$$

Let $\{c_i;\ 1 \leq i \leq N\}$ be defined by (5.5). From (5.6)

$$(5.7) \qquad \log L \leq \sum_{i=1}^{N} \log q(Y_i) - \sum_{i=1}^{N-1} c_i\, q(Y_i).$$

Without the restriction that $q(Y_1) \leq q(Y_2) \leq \cdots \leq q(Y_N)$, the maximum of the right hand side of (5.7) is achieved for $\hat{q}(Y_i) = c_i^{-1}$ for $1 \leq i \leq N$. (For $i = N$, c_i^{-1} is not defined, but the limiting argument used in [10] to obtain (5.3) can be applied to get $c_N = \infty$.) However, $q(Y_1) < q(Y_2) < \cdots < q(Y_N)$ defines a convex set and the right side of (5.7) satisfies Brunk's conditions, so Brunk's result (Corollary 2.1, [3]) can be applied to obtain the maximum at (5.4).

If $U < Y_1$, $Q(U) \geq 0$ and (5.7) reduces to the corresponding statement for $U = 0$, which is maximized by (5.3).

The MLE for $h(y)$ is obtained from (5.3) or (5.4) in the natural way, that is

$$(5.8) \qquad \hat{h}(y) = \hat{q}(y) \exp [-\int_0^y \hat{q}(z)\, dz].$$

The MLE (5.4) can be shown to be a consistent estimator, so that (5.8) is a consistent estimator of $h(y)$.

<u>Theorem 5.2</u>: If $q(y)$ is increasing, then for every t_o

$$q(t_o^-) \leq \lim\inf \hat{q}(t_o) \leq \lim\sup \hat{q}(t_o) \leq q(t_o^+).$$

<u>Proof</u>: The proof follows directly from the consistency of (5.3) (c.f. Theorem 4.1, [10]) after the observation that

$$(N-i)(Y_{i+1} - Y_i) \leq c_i \leq (N-i+1)(Y_{i+1}-Y_i) \quad 1 \leq i \leq N-1.$$

This type of solution has also been obtained for the MLE of a decreasing failure rate density for non-random sample size (c.f. Section 6, [10]).

c. <u>Non-increasing densities</u>, that is the class of densities for which $h(x_1) \geq h(x_2)$ if $x_1 < x_2$. For non-random sample size, Grenander [9] has derived for this case the MLE for a density $h(y)$ and for the corresponding distribution function $H(y)$ (c.f. 3.1, [9]). For an ordered sample $(Y_1, \cdots, Y_n)$ of fixed size the MLE of $H(y)$

is the smallest concave majorant of the empirical distribution function. The MLE for $h(y)$ can be written in the same form as (5.3), namely

$$(5.9) \quad \hat{h}(y) = \begin{cases} \max\limits_{v \geq i+1} \min\limits_{u \leq i} n^{-1}[(v-u)(Y_v - Y_u)], & \text{if } Y_i \leq y \leq Y_{i+1} \\ & (0 \leq i \leq n-1) \\ 0 & \text{if } y > Y_n \end{cases}$$

where Y_o is the left end point of the support of $H(y)$.

For a renewal process with a non-increasing density, a MLE can be obtained within the class $\mathfrak{H} = \{h(x): \int_0^\infty h(x) \leq 1\}$. Let $(X_1, \cdots, X_{N(t)})$ be a sample from a renewal process over $[0,t]$, and let $(Y_1, \cdots, Y_n)$ denote $\{X_i : 1 \leq i \leq N(t) = n\}$ arranged in increasing order. Let $\mathfrak{H}_a$ be the subclass of non-increasing densities $h \in \mathfrak{H}$ which satisfy

$$\int_{Y_{i-1}}^{Y_i} h(y)\, dy = a_i$$

for some fixed constants $a_1, a_2, \cdots, a_n$. For $1 \leq i \leq n$ and $Y_{i-1} < y \leq Y_i$, define

$$(5.10) \qquad h^*(y) = a_i/(Y_i - Y_{i-1}) = h_i$$

and for $y > Y_n$ let $h^*(y)$ be any function which is non-increasing on $[Y_n, \infty)$ and satisfies

$$\int_U^\infty h^*(y)\ dy \le \alpha.$$

For $h \in \mathcal{H}_a$, one has

$$\prod_{i=1}^n h(Y_i)[1 - H_i(U)] \le \alpha \prod_{i=1}^n h_i \equiv \Phi(\alpha, h_1, \cdots, h_n)$$

that is, the maximum of the likelihood function over $\mathcal{H}_a$ is attained at a density of the form (5.10) for some choice of the constants a_i, $1 \le i \le n$. Thus the MLE for $h \in \mathcal{H}$ is obtained by maximizing over all $\mathcal{H}_a \subset \mathcal{H}$ for which the a_i's are non-increasing. Specifically the MLE for $h \in \mathcal{H}$ is that function $\hat{h}$ which maximizes $\Phi(\alpha, h_1, \cdots, h_n)$ subject to

(5.11) $$0 \le \alpha \le 1,\ h_1 \ge h_2 \ge \cdots \ge h_n \ge 0,$$

(5.12) $$\int_0^U h^*(y)dy = 1 - \alpha,\ \int_U^\infty h^*(y)dy \le \alpha.$$

If $Y_{i_o-1} < U \le Y_{i_o}$, $1 \le i_o \le n$, (5.11) can be written as

(5.13) $$\sum_{i=1}^{i_o} h_i(Y_i - Y_{i-1}) + h_{i_o}(U - Y_{i_o-1}) = 1 - \alpha$$

(5.14) $$\sum_{i=i_o+1}^{n} h_i(Y_i - Y_{i-1}) + h_{i_o}(Y_{i_o} - U) + \int_{Y_n}^\infty h^*(x)\ dx \le \alpha.$$

The integral term in (5.14) can be set equal to zero without affecting the likelihood, so (5.14) becomes

$$(5.15) \qquad \sum_{i=i_o+1}^{n} h_i(Y_i-Y_{i-1}) + h_{i_o}(Y_{i_o}-U) \leq \alpha.$$

But $\Phi(\alpha, h_1, \cdots, h_n)$ satisfies Brunk's conditions and (5.11), (5.13), and (5.15) define a convex set, so Brunk's iterative procedure (c.f. Corollary 2.1, [3]) yields the required maximum.

If $U > Y_n$, (5.12) can be written as

$$(5.16) \qquad \sum_{i=1}^{n} h_i(Y_i - Y_{i-1}) + \int_{Y_n}^{U} h^*(y)\, dy = 1 - \alpha$$

and

$$\int_U^{\infty} h^*(y)\, dy \leq \alpha.$$

Pick $h^*(y)$ to be zero for $y > Y_n$. Then (5.16) can be written

$$(5.17) \qquad \sum_{i=1}^{n} h_i(Y_i - Y_{i-1}) = 1 - \alpha, \int_U^{\infty} h^*(y)\, dy = 0.$$

Again (5.11) and (5.17) define a convex set and Brunk's procedure can be applied.

The h^* chosen will possibly have mass at $y = \infty$, which should not be surprising since $U > Y_n$ represents the information that there is an observation larger than all the other observations. The arbitrary character of h^* for $y > Y_n$ results in a similar arbitrariness in the MLE. For $y > Y_n$ the MLE can be extended in any manner which is non-increasing and which maintains the required area.

6. MAXIMUM LIKELIHOOD ESTIMATION FOR A MARKOV RENEWAL PROCESS

Throughout this section, write $N(t) = N$, $u_t = U$, $J_{N(t)} = J$ whenever t is fixed. From (4.2) the likelihood function for a sample function $(J_o, J_1, \cdots, J_N, X_1, \cdots, X_N)$ is

$$L = p_{J_o} [1 - H_J(U)] \prod_{k=o}^{N-1} p_{j_k j_{k+1}} h_{j_k}(X_{k+1}) \tag{6.1}$$

which may be rewritten as

$$L = p_{J_o} \prod_{i=1}^{m} \prod_{k=1}^{m} p_{ik}^{N_{ik}(t)} [1 - H_J(U)] \prod_{i=1}^{m} \prod_{k=1}^{N_i(t)} h_i(X_{ik}). \tag{6.2}$$

Consider a maximum likelihood problem for which the quantities $\{p_{ik}, 1 \leq i, k \leq m\}$ and $\{H_i(x), 1 \leq i \leq m\}$ are not functionally dependent. The likelihood function then factors into two parts given by

$$p_{J_o} \prod_{i=1}^{m} \prod_{k=1}^{m} p_{ik}^{N_{ik}(t)} \tag{6.3}$$

and

$$(6.4) \qquad [1 - H_J(U)] \prod_{i=1}^{m} \prod_{k=1}^{N_i(t)} h_i(X_{ik})$$

which can be separately maximized. If, furthermore, the H_i's themselves are not functionally dependent, then (6.4) can be factored into m parts given by

$$(6.5) \qquad \prod_{k=1}^{N_i(t)} h_i(X_{ik}) \quad , \quad i \neq J_{N(t)}$$

and

$$(6.6) \qquad [1 - H_J(U)] \prod_{k=1}^{N_J(t)} h_J\,(X_{Jk})$$

which can be maximized separately.

Thus the problem of obtaining an MLE for an MRP in which (p_{ij}), $H_1, H_2, \cdots, H_m$ are not functionally related, reduces to three separate maximum likelihood problems: (i) the problem of maximizing (6.3) which is equivalent to finding the MLE $\hat{p}_{ij}$ of the transition matrix of a Markov chain, (ii) the problem of maximizing (6.5) which is equivalent to finding the MLE $\hat{H}_i$ for $m-1$ densities based on non-random sample sizes, (iii) the problem of maximizing (6.6) which is equivalent to finding the MLE $\overline{H}$ of the density of a renewal process. Solutions of problem (i) have been obtained by Billingsley [4]. Problem (ii) is just the classical maximum likelihood problem for which solutions are well known. The solution of problem (iii) has been obtained for a few cases in Chapter 5.

In particular the MLE for an element $Q_{ij}(x)$ of the transition distribution matrix is given by

$$(6.7)\quad \hat{Q}_{ij}(x;t) = \begin{cases} \hat{p}_{ij}(t)\,\overline{H}_i(x;t) & \text{if } i = J_{N(t)} \\ \hat{p}_{ij}(t)\,\hat{H}_i(x;t) & \text{if } i \neq J_{N(t)}. \end{cases}$$

If a functional relation exists between (p_{ij}), $H_1, H_2, \ldots, H_m$ the problem is much more difficult.

REFERENCES

1. Arthur Albert, Estimating the infinitesimal generator of a continuous time, finite state Markov process, Ann. Math. Stat., vol. 33 (1962), pp. 727-753.

2. M. S. Bartlett, The frequency goodness of fit test for probability chains, Proc. Camb. Phil. Soc., vol. 47 (1951), pp. 86-95.

3. H. D. Brunk, On the estimation of parameters restricted by inequalities, Ann. Math. Stat., vol. 29 (1958), pp. 437-454.

4. Patrick Billingsley, Statistical inference for Markov processes, Institute of Mathematical Statistics, University of Chicago Statistical Research Monographs, University of Chicago Press, Chicago, 1961.

5. Patrick Billingsley, Statistical methods in Markov chains, Ann. Math. Stat., vol. 32 (1961), pp. 12-40.

6. Harold Cramér, Mathematical methods of statistics, Princeton University Press, 1946.

7. D. A. Darling, The Kolmogorov-Smirnov, Cramér-von Mises tests, Ann. Math. Stat., vol. 28 (1957), pp. 823-838.

8. Cyrus Derman, Some Asymptotic distribution theory for Markov chains with a denumerable number of states, Biometrika, vol. 43 (1956), pp. 285-294.

9. Ulf Grenander, On the theory of mortality measurement, Part II, Skand. Aktuarietidskr., vol. 39 (1956), pp. 125-153.

10. Albert Marshall and Frank Proschan, Maximum likelihood estimation for distributions with monotone failure rate, to appear in Ann. Math. Stat.

11. Ronald Pyke, Markov renewal processes: definitions and preliminary properties. Ann. Math. Stat., vol. 32 (1961), pp. 1231-1242.

12. Ronald Pyke, Limit theorems for Markov renewal processes, to appear in Ann. Math. Stat., vol. 35 (1964).